DOCUMENT RESUME

ED 179 577 TM 009 858

AUTHOR	Vineberg, Robert; Taylor, John E.
TITLE	An Analysis and Plan of Test Development for the Law Enforcement Basic Training Course.
INSTITUTION	Human Resources Research Organization, Alexandria, Va.
REPORT NO	HumRRO-PP-1-79
PUB DATE	Jun 79
NOTE	23p.
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Achievement Tests; *Behavioral Objectives; *Criterion Referenced Tests; Needs Assessment; Performance Tests; *Police; Role Playing; Simulation; Situational Tests; *Test Construction; Testing Programs; Test Validity; *Training Objectives; *Work Sample Tests
IDENTIFIERS	California

ABSTRACT

A test development plan is described to evaluate police enrolled in the law enforcement basic training course developed by California's Commission on Peace Officer Standards and Training (POST). Some general test methodologies are discussed: performance tests, knowledge tests, and situational tests, including role playing simulations and audiovisual simulations. The test development plan is comprised of three phases: (1) development of test content and strategies; (2) design of diagnostic, proficiency, and certification tests; and (3) test validation. In order to complete phase 1, performance objectives are classified, refined, and given priority ratings; the appropriate strategy for testing each objective is determined; and the weighting for each kind of test is assigned. In phase 2, the three kinds of tests are constructed based on the refined behavioral and knowledge objectives. The three types of tests differ more in terms of purpose and scope, than in context. Test validation procedures are to be developed subsequent to test development. Example of the test instruments to be developed are given. (MH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Professional
Paper
1-79

HumRRO
PP-1-79

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TM

HumRRO

ED179577

TM009 858

An Analysis and Plan of Test Development For the Law Enforcement Basic Training Course

by

Robert Vineberg
John E. Taylor

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
S. LAVISKY

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

HUMAN RESOURCES RESEARCH ORGANIZATION
300 North Washington Street • Alexandria, Virginia 22314

Approved for public release, distribution unlimited.

June 1979

Prefatory Note

In California, the basic training course for law enforcement officers developed by the Commission on Peace Officer Standards and Training (POST) has been adopted by most of that state's law enforcement agencies. As revised in 1973, this is now a performance-based program of instruction consisting of three elements: (1) a detailed listing of approximately 630 performance objectives; (2) some 192 separate instructional unit guides; and (3) a master guide for using these instructional materials in a performance-based instructional system.

The master guide also presents recommendations for evaluating the law enforcement training in terms of evaluation strategies, assessment instruments, and assessment documentation procedures. The Commission on Peace Officer Standards and Training recommends three general classes of tests: diagnostic, proficiency, and certification.

In that connection, the Commission contracted with HumRRO to conduct an analysis and to produce a test development plan for the basic law enforcement course. Although HumRRO's final report was specific to the Commission's needs, the general applicability of the prescription was such that the Commission granted HumRRO permission to publish the report in its Professional Paper series.

TABLE OF CONTENTS

	Page
INTRODUCTION	1
Need for Tests	1
POST's Test Requirements	2
CHARACTERISTICS OF TEST METHODOLOGIES RELEVANT TO POST'S NEEDS	2
Performance Tests	2
Knowledge Tests	3
Situational Tests	4
Role Playing Simulations	4
A–V Simulations	5
TEST DEVELOPMENT PLAN	6
Development of Test Content and Strategies	6
Determine Priorities and Amplify Performance Objectives	6
Identify Testing Strategy for Each Objective	8
Determining Weighting to be Given Each Strategy	9
Design and Construction of Diagnostic, Proficiency, and Certification Tests	10
Develop Sampling Plan and Test Designs	10
Construct Tests	11
Test Validation	11
EXAMPLES OF TEST INSTRUMENTS TO BE DEVELOPED	12
Example of a Performance Test	12
Examples of Knowledge Tests	14
Example of A–V Simulation Test	15
RECOMMENDED DELIVERABLES	18
RECOMMENDED SCHEDULE OF ACTIVITIES AND MANPOWER LOADING REQUIREMENTS	19
STAFFING REQUIREMENTS	19

An Analysis and Plan of Test Development for The Law Enforcement Basic Training Course

ROBERT VINEBERG and JOHN E. TAYLOR

Need for Tests

POST's Basic Course, which was first established in 1960 and revised in 1964, has been adopted nearly universally by California's law enforcement agencies. Recognizing the need to reevaluate, update, and improve the course's training standards, a consortium including the POST Commission, local law enforcement agencies, and community colleges undertook a cooperative effort to modify the course in 1973. That effort, now completed, produced a revised basic course training system. It is organized into three basic components documented by the following publications.

- *Performance Objectives for the POST Basic Course*, a detailed listing of the approximately 630 objectives which constitute the new standards for basic training.
- *Instructional Unit Guides*, 192 separate unit guides designed to serve as model units of instruction for the course instructor and training manager.
- *Management Guide*, the master guide for implementing performance-based instruction and using the instructional unit guides.

At this time, selected basic academies are undertaking implementation of the revised course and are beginning a gradual conversion of their basic course from conventional techniques to the proposed performance-based system. The materials and guidance provided in the three documents cited above (objectives, guides, and management) clearly specify the instructional content, instructional techniques, and overall management system to be employed in training the basic police officer. They also provide guidance in the fourth major basic component of the course, evaluation of performance-based training. Pages 13 through 20 of the Management Guide provide general guidance as to the development of evaluation strategies, assessment instruments, and assessment documentation procedures. Discussion and examples of types of tests, response categories, and test items are provided as guidance to the course manager and instructional staff to develop test items and assessment instruments from the statements of performance objectives. Further, each unit guide contains evaluation examples sufficient to minimally meet course needs. Additional sophisticated and more complete tests remain to be developed.

When the necessary additional tests and supporting documentation have been developed, POST will have the means for insuring that training by the various academies will be directed to achieving the same ends. Standardization of test items and criteria of success will foster adherence to the same standards and eliminate the ambiguity and varying interpretations of instructional objectives that may result from each academy's building its own tests. The academies will have the necessary tests for assessing their students, and POST will have the means of maintaining quality control of the instructional system.

POST's Test Requirements

POST has requested that three general classes of tests be developed for use in conjunction with the basic course: Diagnostic, Proficiency, and Certification. Briefly, they would be used as follows.

Diagnostic - This test, more likely a series of tests, would be administered to the basic recruit upon his entry into the basic course. Test results would be used by course personnel to diagnose his specific strengths and weaknesses vis a vis the basic course objectives. They would be used to design his unique path for progress through the course, indicating which objectives he already has learned and those in which he must receive training. The results would indicate where in the 12 course functional areas he should undergo instruction in order to meet the course graduation standards.

Proficiency - This test(s) would be used by course personnel to assess the student's mastery of the course content as he progresses through the course. Such tests would be developed for use after training in each functional area, and would be designed to measure the students' performance of that functional area's major objectives.

Certification - This test(s) would be used by POST personnel to ensure that course graduates demonstrate the standards of proficiency established for the course. These tests might be as long as two days in length, sampling the graduate's performance in all the functional areas. The basic certification tests would be general, and would assess performances required across all regions of the state. These certification tests will serve as one of POST's vehicles for maintaining quality control.

There is a distinct possibility that the three classes of tests may be alternate forms of a common test bed, and that they may not have to necessarily assess qualitatively differing knowledge and skill domains.

CHARACTERISTICS OF TEST METHODOLOGIES RELEVANT TO POST'S NEEDS

The problem of diagnostic and criterion measurement in training in its simplest form is the identification of tests that are (a) relevant to actual job behavior, (b) standardized and reliable in their measurement, (c) economically feasible to administer, and (d) providing of sufficient coverage to be representative of the training or job. In the next section we will suggest a plan for developing tests to be used in conjunction with the POST basic course to meet these requirements. As background to this plan, it is useful first to discuss the characteristics of two general test methodologies—performance tests and knowledge tests; and two varieties of situational tests, role playing simulations and video (A—V) simulations—as they relate to peace officer training.

Performance Tests

Performance tests, or job-sample tests, as they are sometimes called, measure behavior that is the same as or very similar to that which occurs in the job situation. Likewise, the environmental conditions or cues that the performer responds to are similar to those encountered in actual job performance.

Performance tests are necessary when *skilled* behavior is to be evaluated. A convenient way to decide whether or not behavior involves skill is to consider how much practice is required to become proficient. If a person can read about or be told about a task he has never performed and then do it with little or no practice, the task can be considered non-skilled or low in its skill requirements. If practice is necessary, then the task is skilled and a performance test is required to evaluate it. This is because the "practice" yardstick has

demonstrated that a capability cannot safely be inferred or assumed solely on the basis of a verbal description of performance.

Some useful cues to identifying skill are behaviors that:

- Involve finely tuned acts of physical coordination.
- Require quick reaction.
- Are conceptual rather than procedural.
- Involve time-sharing—doing two things at once.
- Involve interactions with other individuals or groups of people.

Performance tests that are used to evaluate skilled behavior inevitably involve some degree of simulation. While the cues that are presented and/or the responses that are required may be somewhat degraded, their form or modality is maintained. For example, since a policeman who is directing traffic responds to moving vehicles, a performance test of traffic control requires some representation of vehicles that are actually moving. In a simulation, an examinee, rather than seeing actual traffic directly, might respond to traffic on a film or video tape. In a simulation of firing against live targets, a laser weapon might be substituted for an actual rifle. The officer still tracks the target and squeezes a trigger, even though his total response has changed somewhat in that he no longer has to be set to absorb a recoil.

The key requirement of a performance test is that *critical* cues and responses be represented with a degree of fidelity that will permit generalization from the simulated test situation to the actual job situation. Unfortunately, no clear rules exist for judging the extent to which physical modifications of tasks affect their predictive validity. Rational appraisals and prior experience are, at present, the best basis for decisions.

The maintenance of behavioral or psychological fidelity presents an even more difficult problem when the task to be simulated involves not merely perceptual, motor, or cognitive skill, but requires interactions among persons and demands some degree of *interpersonal skill*. Interpersonal skill is, of course, required in many of the tasks performed by peace officers.

Finally, performance tests are usually quite expensive. They often are costly in terms of (a) time required for training of scorers, (b) time needed for individualized administration of the test itself, and (c) the cost of equipment and physical resources they require. Because performance tests are generally expensive, they are *inefficient* to use in situations where they are not required.

- Performance tests *should* be used where skilled performance is to be evaluated and where simulations with adequate fidelity can be provided.
- Performance tests *should not* be used where:
  1. Knowledge rather than performance is to be evaluated.
  2. Non-skilled performance is to be evaluated.
  3. Skilled performance is to be evaluated, but simulations with adequate fidelity cannot be provided.

Knowledge Tests

Knowledge tests measure information that mediates job behavior. Usually a person is required to recall or recognize information that describes how a particular task is to be performed or information that provides for adjustment to specific situational demands in performance.

In knowledge tests, the cues or conditions that are responded to may be quite similar to those encountered in actual job performance or they may be very different. The primary characteristic of a knowledge test is that the responses to be made are verbal or symbolic and generally are different from the behavioral responses made in actual task performance (or on a performance test.)*

*Performance tests for tasks that require the manipulation of information or symbolic material (e.g., filling out forms or solving mathematical problems) will also require verbal or symbolic responses much the same as knowledge tests.

- Knowledge tests cannot be used to evaluate skilled behavior, but they do have a variety of virtues:
  a. They are economical to administer.
  b. They can assess information about many more tasks than can be covered with performance tests in a finite amount of time.
  c. Administration and socring are more readily standardized than with performance tests; they tend to have better psychometric properties.
- Knowledge tests *should* be used where information alone is to be evaluated when:
  1. Performance is non-skilled.
  2. Performance is skilled, but cannot be simulated with adequate fidelity and a test of knowledge—though not sufficient to predict performance—can be useful for screening or diagnostic purposes.
- Knowledge tests *should not* be used where skilled performance is to be evaluated and can be tested directly or adequately simulated.

Situational Tests

A situational test is one that places a person in a situation closely resembling or simulating a "real life" circumstance. Generally, these tests are not used to measure technical abilities per se, but rather, judgmental and decision making skills; interpersonal skills; and emotional, attitudinal, motivational, and other personality variables. Situational tests are often used in an attempt to determine *what a person will do* or *what he will observe and recommend* in a particular situation without artificially cueing him to perform. Where the focus is on how a person acts (response) these tests can be considered a particular variety of performance or job-sample test. Where the focus is on an analysis of the situation (stimulus) and the prescription of a solution, these tests can be considered a form of knowledge test. Some situational tests will, of course, focus on both the stimulus and the response. Two particular varieties of situational tests concern us: role playing simulations, where the focus is often on how a person *responds* to particular situational events; and A—V simulations where the focus is on a person's *understanding and analysis* of a situation.

Role Playing Simulations

Role playing simulations, that concentrate on a person's response to situational requirements, pose two major difficulties:

1. Providing for adequate *fidelity* of simulation so the behaviors that emerge can be considered a valid representation of the performance that would be exhibited in response to the demands of the real situation.
2. Maintaining *standardization* of conditions so that measurement of the response will be reliable.

For example, it is extremely difficult to duplicate the stress that occurs and can affect a peace officer's response in dealing with a family dispute situation, or a racial situation, or when he is subject to physical attack. No matter how effectively "role players" or "plants" are trained, the real life impact of the emotional component in these situations is practically impossible to achieve. At the very least, the officer knows he is performing in relation to people who are acting and who are observing and possibly evaluating him. Thus:

1. He has considerable reason to modify his behavior from what he might actually do, to that which appears most advantageous to him.
2. He is able to *select* his behavior since the critical conditions (e.g., stress) that might determine his real life actions are absent.

In tasks involving *interpersonal* behavior, another person or group of people represent an important part of the environment to be controlled (that is, standardized) from one role playing situation to another, and people are notoriously difficult to *standardize*. For example, in simulations where an attacker armed with a knife is to be disarmed, or a resisting or drunken suspect is to be controlled, different actors will inevitably use

movements that necessitate different coping responses on the part of the performer and that present him with tasks of varying difficulty.

The extent to which issues of *fidelity* and *standardization* are critical depend upon the purpose of the role playing simulation. Let us consider the use of role playing for three different purposes:

1. To predict how persons are likely to perform sometime in the future.

2. To provide practical exercises in training them.

3. As criteria for assessing their actual performance in a job.

When role playing is used as a predictor of future performance, it is desireable to measure samples of behavior that are as similar to the criterion as possible. The more consistency there is between pre-employment samples of behavior or samples of behavior taken in training, and relevant dimensions of job behavior, the more likely the test will have predictive validity. While as much overlap as possible is obviously desirable, total fidelity is usually not essential. Samples of behavior in role playing situations that employ at least some of the dimensions of actual job behavior should provide better prediction than other kinds of tests and inventories.* Since whether or not a satisfactory level of prediction is achieved is an empirical matter, some departure from complete fidelity can be tolerated until the evidence on the actual predictive value of the test, is in. Standardization remains a requirement as it always does when a test is to be used for purposes of *measurement*.

When role playing is used for training purposes—its most popular use—as much realism (fidelity) as possible is desireable, but standardization is avoided. A necessary requirement for the learning of diagnostic, decision making, and human relations skills is the opportunity for the trainee to experience constantly changing (non-standardized) interactions with others. Here, he experiments, makes errors, replays roles, and becomes aware of the implications of his actions and the actions of others for him. Spontaneity and variation, rather than standardization, is the way in which fidelity is achieved in role playing and is the key to the acquisition of interpersonal skill.

When actual job proficiency is to be measured—the goal in the development of tests for the POST basic course—the need for fidelity and standardization is maximal. However, *neither role playing nor any other testing strategy* is adequate to fully assess interpersonal and related skills. The most that can be concluded is that the person being tested is proficient in a role playing situation that has many, but not all, of the charactertistics of the real life situation. Skills and behaviors that are *necessary*, but not *sufficient* for job proficiency will have been demonstrated.

Because role playing as a test methodology (1) is not completely satisfactory to measure interpersonal skill, (2) is beset with problems in standardizing the behavior of supporting role players, and (3) is costly, its "cost-effectiveness" or value as a criterion measure in each specific situation must always be reckoned. As with simulations in general, there are no clear rules for making this judgement. Estimates must be based on the practical importance of the bahavior to be measured, the adequacy of the simulation and standardization, and the overall cost.

Before leaving the topic of role playing as a particular variety of situational test, its use as a means to avoid artificial cueing should be mentioned again. Occassionally, it is desireable to use role playing not to assess interpersonal behavior, but to determine whether a person will perceive different or unrelated requirements in a situation and will initiate behavior on his own. In a procedure sometimes called *functionally integrated testing*, tasks or events are embedded or grouped together in larger segments of job behavior. The performer is given a general instruction, such as to patrol a problem area. He must *recognize and respond* to all of the relevant aspects of the situation without being specifically told to do so. This procedure is particularly suited to evaluating an individual's observational, cautionary, and other responses to situational demands and subtleties.

A—V Simulations

A—V simulations concentrate on representation of the stimulus aspects of a situation and are most

*Wernimont, P.F. and J.P. Campbell. "Signs, Samples, and Criteria", *Journal of Applied Psychology*, Vol. 52 (1968).

useful for evaluating how adequately a person understands its requirements. As mentioned earlier, they focus on maintaining fidelity with respect to what a person might observe in a situation, rather than what he might do. When used for testing, A—V simulations should be considered tests of knowledge, rather than tests of performance and should be used wherever knowledge testing is appropriate.*

A—V simulation tests can employ either of two general formats:

1. A situation is depicted and the person being tested indicates what should be done.

or

2. A person is depicted performing in a situation and the person being tested critiques the performance.

An advantage of the latter procedure is its capacity to provide for examination of an entire sequence of task performance. The entire performance, with all of the detail that only performance can provide, can be used as the test stimulus. The viewer must consider each aspect of the performance portrayed and continuously judge whether what he is viewing is being done correctly (or incorrectly).

These approaches are somewhat analogous to traditional paper and pencil tests of recall and recognition. However, in either A—V simulation format, provision can be made for an objective response; e.g., multiple choice answers.

A—V simulation testing can be contrasted with standard paper and pencil tests and oral "talk throughs" or written descriptions of performance that are open-ended. Standard tests can deal with only selected aspects or elements of performance; i.e., they require abridgement of performance and its description in verbal form. Open-ended written tests are plagued by serious standardization problems in that they place a heavy burden not only upon the examinee's verbal skills, but upon the test administrator-scorer's ability to accurately interpret the information he is given in order to classify it as "right" or "wrong".

Issues to which careful attention must be given in designing A—V simulators include the following:

- Realistic selection of performance errors to be introduced into the scenario.
- Provision for introducing errors that would ordinarily preclude the continuation of task performance.
- Provision for objective scoring while minimizing the articicial dueing of answers.

TEST DEVELOPMENT PLAN

The development of tests for use in conjunction with the POST basic course should occur in three phases.

- Development of Test Content & Strategies
- Design of Diagnostic, Proficiency, & Certification Tests
- Test Validation

Development of Test Content & Strategies

Determine Priorities & Amplify Performance Objectives

The majority of performance objectives given in *Performance Objectives for the POST Basic Course*, The Commission on Peace Officer Standards and Training, January 1977, call for information of varying degrees of specificity to be identified, understood, or otherwise developed by peace officer students. Some objectives call for a display of behavior. As a first step in development, the objectives should be classified and refined as follows.

a. *Specify the level of importance and priority of each objective for functioning in the job.* Particularly important or crucial objectives may be identified as those which:

1. Can have particularly serious consequences when mastery does not occur (e.g., threat to life or personal harm).
2. Are common sources of failure in job performance.

*Occasionally fidelity of stimulus and response are both maintained, as in driver training simulators. Testing with these simulators are properly considered performance tests.

Importance should be determined with the assistance of panels of police experts. This information would be used as part of the criteria for selecting content for different types of tests (i.e., Diagnostic, Proficiency, and Certification) and when time and cost considerations dictate, the sampling of objectives within a particular type of test.

b. *Determine whether some knowledge items should be replaced or augmented by a requirement for a display of behavior.* Some of the objectives presently require a display of behavior in a demonstration or practical exercise (e.g., Vehicle Control objectives, 6.6.0; Pedestrian Approach objectives, 8.7.4; Traffic Control objectives, 9.12.1 - 9.12.3). Appropriately, most of the objectives require a display of knowledge as evidenced by providing information (e.g., Probable Cause objectives, 3.6.1; Concepts of Evidence objectives, 4.1.0) or by analyzing a situation (e.g., Obstructions of Justice Law objectives, 3.8.0; Seizure Concepts objectives, 4.8.1). A few objectives require a display of knowledge where more valid measurement may be obtained by requiring a display of behavior (e.g., Patrol "Hazards" objectives, 8.6.1 and 8.6.2). Each knowledge objective should be examined to determine whether translation to a behavior objective is appropriate and feasible.

c. *Specify, as needed, the specific elements of knowledge or behavior that define correct performance and will facilitate standardization and objectivity in scoring.* Some of the objectives are probably not specified in sufficient detail to permit an unambiguous determination of mastery. For example, Community Service Concept objective, 2.1.2, states: "Given a role playing situation encompassing any one of the police community service responsibilities identified in Objective 2.1.1 [Order Maintenance, Crime Prevention, Public Education, Delivery of Service, Enforcement of Law] the student will react in a manner that will promote positive police-community service, as noted on a POST approved, predesigned rating form."

The POST approved rating from provided in instructional unit 2.1.0 does not specify any of the characteristics that define a "positive policy-community service" (see following).

EVALUATION EXAMPLES:

POST EVALUATION SHEET
2.1.0

Community Service Concept _____

YES	NO	
☐	☐	Was student aware of the service he was providing and so noted to the instructor at the end of role playing.
☐	☐	Did the student action provide a positive community service with the public.

COMMENTS _____

_____

Search Concepts objective, 4.7.4, indicates that knowledge is to be demonstrated.

> The student will explain the conditions under which at least 4 of the following types of legally authorized searches may be made: (Case Law Decisions)
>
> A. Consent searches
> B. Search warrant
> C. Plain sight
> D. Incidental to arrest
> E. Probable cause search

However, the objective as stated, does not indicate the critical element of information that should appear in a correct explanation of each condition. For example, the plain sight doctrine indicates that *evidence may be seized if it is observed from a position where an officer has a lawful right to be.* While some may question the need for such a degree of specificity in an objective, experience has shown that it is necessary if standardization of scoring is to be maintained when the objective is translated into a test item.

Each behavior objective and each knowledge objective should be defined sufficiently so that two or more persons observing the same performance will be consistent in what they score as correct or incorrect. Behavior objectives should be defined in terms of the presence or absence of particular behaviors or particular characteristics of products. Knowledge objectives should be defined in terms of specific concepts or elements of information. When criteria are ambiguous, it is generally not worthwhile to assess performance. Classifying persons as competent or incompetent on the basis of tests that are known to have imprecise criteria is far worse than not testing at all.

d. *Identify key elements of behavior for behavior objectives.* Key elements of behavior are those that have been identified as:

1. Having serious consequences of failure.
2. Being common sources of failure in the performance of a task.

Similar to classifying objectives as to importance, here the individual elements or components of behavior objectives are analyzed. Not all behavior objectives will have key elements. The significance of key elements arises when performance tests of lengthy or time consuming tasks are to be constructed and it is desireable to reproduce only part of the required behavior in the test situation. Also, it is often necessary to score only the key elements in a sequence of behaviors.

Identify Testing Strategy for Each Objective

After the performance objectives have been refined for test construction, the appropriate strategy for testing each objective should be determined. In keeping with our earlier discussion:

- Behavior objectives that require the display of skill should be tested by means of conventional performance tests if adequate fidelity can be provided. If adequate fidelity cannot be provided, this should be acknowledged. In such instances, a partial criterion can be obtained by using knowledge tests to evaluate the informational component in a task. The results of such testing can be used for screening or diagnostic purposes, but cannot be used to demonstrate conclusively that proficiency has been attained.
- Behavior objectives that require the display of interpersonal skill cannot be totally assessed. Short of this, an understanding (knowledge) of requirements in interpersonal situations can be tested by means of conventional knowledge tests, role playing simulations, and A—V simulations. From a testing perspective, A—V simulations will in general be the most desirable choice.

- Behavior objectives that do not involve skill shouuld be translated into knowledge objectives and tested with either conventional paper and pencil tests or with A—V simulations.
- Knowledge objectives should be tested with either conventional paper and pencil tests or with A—V simulations.

Determining Weighting to be Given Each Strategy

To develop a preliminary estimate of the weighting to be given to different testing strategies during the development, *Performance Objectives for the POST Basic Course, January 1977,* were analyzed. The results of this analysis are shown in Table 1.

Approximately 65 percent of the objectives could be evaluated by means of conventional knowledge tests, 15 percent by means of performance tests, five percent by A—V simulations, and 15 percent by knowledge tests or A—V simulations. A more precise specification of the weighting to be accorded the different strategies depends upon the characteristics of the objectives that emerge following refinement; i.e., objectives that are identified as highly important, knowledge objectives that should be restated to require a display of behavior, and objectives that are not tested because they involve a skill component for which testing is not feasible.

It is anticipated that the weighting of performance test items will increase somewhat after refinement of the objectives. Many behavior objectives (requiring performance testing) are located in the Force and Weaponry, and in the Patrol Procedures Functional Areas, where they are perhaps more likely to be assigned a high priority for testing.

TABLE 1
Frequency of Different Test Strategies Judged to be Appropriate for Objectives of the POST Basic Course

FUNCTIONAL AREA	TEST STRATEGY					
	Knowledge	Performance	A—V Simulation	Knowledge or A—V Simulation	Performance or A—V Simulation	
Professional Orientation	31	—	—	—	—	
Police Community Reelations	20	11	11	—	—	
Law	60	1	—	76	—	
Laws of Evidence	29	—	—	—	—	
Communications	12	7	—	—	—	
Vehicle Operations	21	12	—	—	—	
Force & Weaponry	21	23	—	5	—	
Patrol Procedures	124	20	6	4	—	
Traffic	31	13	12	—	—	
Criminal Investigation	18	5	—	1	4	
Custody	27	—	—	1	—	
Physical Fitness & Defense Techniques	16	13	1	—	—	
TOTAL	410	94	25	87	4	(620)*
PERCENT	66	15	4	14	1	

*10 objectives from the Criminal Investigation Area that Focus on Local Agency Procedures have been Omitted.

Design & Construction of Diagnositc, Proficiency, and Certification Tests

The refined behavior and knowledge objectives, and the strategies that have been selected for evaluating them provide the basic content for the construction of Diagnostic, Proficiency, and Certification tests for use in conjunction with the POST basic course. In most instances, these three different types of tests will be quite similar. They are differentiated largely in terms of purpose and scope, rather than in terms of their content per se.

Diagnostic Tests are intended to be used for making placement decisions within the basic course at a police academy. They would be designed to identify those instructional units in which training is needed for persons entering an academy. To serve this function, diagnostic tests would provide the broadest coverage of the three types of tests and probably would cover all performance objectives. Because they are to be used solely for placement, rather than qualification or certification, conceivably, they might concentrate more heavily on knowledge requirements.

Proficiency Tests are intended to be used by the basic academies as within and end-of-course tests to establish that objectives required for graduation have been met. In addition to objectives of the POST basic course, they may ultimately include objectives that have been established locally to cover specific regional requirements. Proficiency tests should assess knowledge and skill.

Certification Tests are intended to parallel the proficiency tests, but are to be administered under state auspices. Their primary purpose is to assure quality control of the graduates of local academies. With the exception of locally established objectives, they will be virtually identical to proficiency tests. They also should assess both knowledge and skill.

Develop Sampling Plan & Test Designs

Since most jobs consist of a multitude of requirements, sampling from among the elements that define a job is almost always a step that must be taken before any actual test construction begins. In the present instance, it is unlikely that it will be possible to test all performance objectives in each of the three varieties of tests. In particular, performance testing and A—V simulation testing is likely to make this prohibitively expensive. When the population of refined performance objectives and their associated test strategies have been determined, procedures for sampling objectives should be developed in conjunction with POST. Dimensions to be treated in the sampling plan will be skill type (e.g., interpersonal, perceptual, motor), knowledge domain (instructional unit), and importance. It is possible that objectives with a high priority will not be sampled, but will be represented completely in the proficiency and certification tests.

An additional step related to the sampling or content selection process is the development of test designs (scenarios) for A—V simulations and performance tests. When individual objectives to be tested have been identified, they should be examined together. This will often suggest ways to combine situations to promote efficiency, and perhaps more importantly, provide for greater fidelity or realism. Such functionally integrated testing permits avoidance of artificial cueing that often occurs when a single objective is to be evaluated. Review of the objectives for purposes of functional integration will usually lead to some modification of the sample that has been selected for testing. In any modification the representation of objectives with similar characteristics should, of course, be maintained in the same proportion as that dictated by the original sampling plan.

Construct Tests

Test construction should include the following:

Performance Tests

- Identification and specification of test facilities, equipment and material, and resource material requirements.
- Development of training materials and instructions for test administrators.
- Preparation of instructions for examinees.
- Development of score sheets and scoring procedures.

Conventional Knowledge Tests

- Writing of test questions.
- Preparation of answer sheets and scoring and processing instructions.
- Preparation of test administration and examinee instructions.

A—V Simulations

- Development of scenarios.
- Design of testing and performance critique procedure.*
- Development of answer sheets and scoring procedures.
- Preparation of test administration and examinee instructions.

- Training of actors/performers.**
- Production of A—V tapes.**

Test Validation

Simply speaking, validity deals with the question—Do test scores tell us what we want to know? For our purposes, do the scores tell the truth about a peace officer's attainment of the behavior objectives and knowledge objectives that have been specified for the POST basic course? This is, is the knowledge or behavior called for in a test an adequate representation of the knowledge or behavior specified in the objectives? Because the objectives are to be stated in terms of very specific knowledge and behavioral requirements, this issue is fairly trivial. It is particularly so where *all* objectives are to be tested and the adequacy of sampling is not in question.

Since the objectives and test items will be virtually synonymus, the real question becomes, how good are the objectives? Do they and their associated training adequately prepare a peace officer to perform his job? This question can only be answered when trainees have graduated from the POST basic course and performed in the job so their performance can be observed and evaluated. Procedures to be used to evaluate job performance and validate the objectives-training-testing system could be developed subsequently, and would be a logical follow-on to the test development effort.

*Where A—V simulations are used to display behavior to be critiqued by the examinee, a particular problem arises. The instructions given to the examinee must strike a balance as to level of detail. He must be given enough information to know how to organize his analysis of the performance he is viewing and how to record what he believes to be correct and incorrect behavior. If he is given too little guidance, the test becomes open-ended and therefore, unstandardized. If too much detail is given, he is virtually given a checklist for following and scoring the performance. Such over-cueing can obviously transform the test into a measure of the examinee's ability to read and to interpret explanations of task performance.

**To provide standardization the training of actors and the actual production of A—V tapes for *certification tests* should be undertaken by POST rather than by the basic academies. Estimated costs for developing tests have not included costs associated with the actual production of A—V simulations. See page 35.

EXAMPLES OF TEST INSTRUMENTS TO BE DEVELOPED

Example of a Performance Test

The following example was derived from consideration of Learning Goal 9.12.0, Traffic Control, and its associated performance objectives. This example is not to be interpreted literally as *the* test recommended for assessing an officer's ability to direct traffic. It is presented only as an example of how such a performance test should be constructed and what actions might be called for and assessed. If appropriate, in an actual test the correct hand and arm movements might also be specified. In order to flesh out a complete example, considerable information was assumed beyond that provided by statements 9.12.0 through 9.12.3.

PERFORM POINT CONTROL OF TRAFFIC (AT 4-WAY INTERSECTION)

EQUIPMENT

Standard police uniform and equipment.

Reflectorized crossbelt, vest, batons, flashlight, or other special equipment used by agency in traffic control.

INSTRUCTIONS TO TESTER

This is a test to see how well the student can perform point control of traffic under condition of day and night visibility. Choose an intersection of two streets with heavy traffic. Give individual being tested all information concerning the traffic control requirements at the intersection and the special equipment to be used.

TEST SITUATION

You are assigned to perform traffic control duty. Your traffic control point is located at the intersection of two major streets at _____ and _____. You will have _____ minutes to complete this test.

TEST STANDARDS

_____ percent of the performance measures must be completed correctly.

GIVE STUDENT ALL INFORMATION CONCERNING THE CONTROL REQUIREMENTS

AT THE INTERSECTION, THE LOCATION OF THE INTERSECTION, AND THE SPECIAL EQUIPT.

	Test #1		Test #2		Test #3	
	Pass	Fail	Pass	Fail	Pass	Fail
A. Day visibility conditions:						
1. Assumes proper stance for directing traffic.						
2. Directs traffic from right to proceed straight ahead.						
3. Directs traffic from left to proceed straight ahead.						
4. Stops traffic from right.						
5. Stops traffic from left.						
6. Stops traffic from front.						
7. Stops traffic from rear.						
8. Directs traffic from right to turn right.						
9. Directs traffic from left to turn right.						
10. Directs traffic from right to turn left.						
11. Directs traffic from left to turn left.						
12. Varies traffic flow to eliminate jam ups.						
B. Night visibility conditions:						
1. Dons reflecting clothing and assumes proper stance for directing traffic, using batons and/or flashlight.						
2. Directs traffic from right to proceed straight ahead.						
3. Directs traffic from left to proceed straight ahead.						
4. Stops traffic from right.						
5. Stops traffic from left.						
6. Stops traffic from front.						
7. Stops traffic from rear.						
8. Directs traffic from right to turn right.						
9. Directs traffic from left to turn right.						
10. Directs traffic from right to turn left.						
11. Directs traffic from left to turn left.						
12. Varies Traffic flow to eliminate jam ups.						
C. Maintains traffic flow through intersection.						
D. Executes signals with distinct, complete motions.						
E. Uses whistle to augment hand and arm signals.						
F. Maintains control.						
G. Demonstrates professionalism, good community relations, and correct police image.						

Examples of Knowledge Tests

The two following examples illustrate (1) a test emphasizing recall (the first three of four items), and (2) a test emphasizing recognition (the last four of five items). The "recall" test was derived from consideration of Learing Goal 7.3.0, Deadly Force, and its associated performance objectives; the "recognition" test was derived from consideration of Learning Goal 3.6.0, Probable Cause, and its associated performance objectives. Again, these examples are not to be construed as *the* tests being recommended; they are illustrations of a testing strategy, not detailed test content.

TEST EMPHASIZING RECALL

DETERMINING WHEN TO USE DEADLY FORCE

1. When determining whether to use deadly force, an officer must consider:

   _____

   _____

2. Name four circumstances where the use of deadly force would be justified.

   _____

   _____

   _____

   _____

3. Deadly force may be used when it reasonably appears to be necessary to prevent the commission of a serious offense against persons that involves violence and threatens death or serious bodily harm, such as:

   _____

   _____

4. Indicate whether deadly force is justified in the following situations:

	YES	NO
Self defense		
Prisoner is escaping		
Individual is resisting arrest		
Rape is in progress		
Arson is in progress		
Individual is setting a bomb		
Armed robbery is in progress		
Burglary is in progress		
Knifing is in progress		
Police officer is being beaten		

TEST EMPHASIZING RECOGNITION

DETERMINING PROBABLE CAUSE

1. Prior to making a legal apprehension, you must determine probable cause. Write the legal definition of "probable cause" in the space provided.

   _____

   _____

2. You have just apprehended a man for aggravated assault. Do you have probable cause to search him?

   _____ Yes _____ No

3. You arrive at a scene where a robbery has just taken place. You observe an individual running from the scene. Do you have probable cause to apprehend him?

   _____ Yes _____ No

4. You have been dispatched to a residence to investigate a complaint of destruction of private property. You arrive and the owner of a vehicle with the windshield broken out points to a person and states, "I saw him throw a rock through the windshield of my car." Do you have probable cause to apprehend him?

   _____ Yes _____ No.

5. You stop and apprehend a known felon in a vehicle. After you have placed him in the vehicle you ask him if you may search his vehicle. He says "yes" and during the search you find a bag of morphine under the seat. Is this a legal search?

   _____ Yes _____ No

Example of A—V Simulation Test

This example was derived from consideration of Learning Goals 9.13.0, Traffic Accident Investigation, and 9.14.0, Traffic Accident Field Problem. As with the preceding examples, it is intended only to illustrate how such a test might be constructed, and the major phases that might be included in the test scenario. It is not to be interpreted literally as the test for assessing objectives 9.13.1 through 9.13.5 and 9.14.1. A test scenario providing detailed visual and audio scripts has not been included because of our limited knowledge of the specific step-by-step police procedures employed in traffic-accident investigation. Such detailed scripts would have to be written in close coordination with subject matter experts, versed in traffic accident investigation procedures, to ensure that the details of correct procedures are accurately reflected in the test.

A—V SIMULATION

TRAFFIC ACCIDENT INVESTIGATION

Instructions to Tester

Provide the individual being tested with the recording form this test and a pencil for recording his notes as the A—V simulation is played. To ensure standardization of the testing situation read the following instructions to the individual. "You are about to view a A—V simulation of an officer conducting a traffic accident investigation. You will view all phases of the investigation beginning with the officer's being dispatched to the scene, and endingwith the clearing of the accident scene. You will have a clear view of everything the officer does, and you will hear all conversations that occur.

This simulation has been designed to determine how well you can differentiate good from poor performance. You will observe that this investigating officer does perform some of the essential steps of a proper accident investigation, and that he omits others. You have been provided with a two-column recording form on which you are to make notes as you follow the officer's progress through this investigation. Under *STEPS PERFORMED* indicate the essential steps the officer included in his investigation; under *STEPS OMITTED* indicate the essential steps the officer failed to include in his investigation.

Make your entries in each column as the simulation proceeds. At the end of the simulation check your notes for completeness. You should have a series of entries under the *STEPS PERFORMED* column, and a series of entries under the *STEPS OMITTED* column. When you are finished turn your completed form in to me."

After reading these instructions and answering any questions, play the test tape. At the conclusion of the tape allow the individual time to check his recorded entries. When he is finished, collect his completed form and check it against your checklist.

(The materials presented below indicate the areas in which detailed visual and audio scripts would be developed.)

VISUAL	AUDIO
Scene 1	
Police vehicle driving down highway; vehicle speeds up; emergency lights and siren go on.	Conversation between dispatcher and responding officer concerning accident and additional information being requested by officer.
Scene 2	
Scene of accident. Officer arrives, takes charge, surveys the situation and looks after the injured.	Conversations among officer, drivers of cars involved and injured parties as he determines severity of accident and extent of injuries.
Scene 3	
Officer on radio at police vehicle.	Conversation between officer and dispatcher as officer reports accident situation and requests ambulance and wrecker.
Scene 4	
Ambulance arrives, loads injured and leaves.	Conversation among officer, ambulance driver and injured parties as officer obtains identification, addresses, destination of ambulance,. . .

Scene 5	
Other cars stop at scene, then move on.	Conversation between officer and other drivers as they offer help, he requests they move on, controls situation.
Scene 6	
Officer interviewing drivers of car involved in accident.	Conversation between officer and drivers as he obtains identification, interviews each, and gathers information about how the accident occurred.
Scene 7	
Officer inspecting accident area.	Narrator indicates that officer is inspecting area, making measurements,. . .
Scene 8	
Wrecker arrives.	Conversation between officer and wrecker driver.
Scene 9	
Wrecker towing car away.	Conversation of officer with drivers of cars as he ascertains their near-future planned whereabouts, dismisses them, checks his notes, and clears scene of accidents.

ILLUSTRATION OF PARTIAL CHECKLIST FOR SCORING INDIVIDUAL'S RECORDING OF THE STEPS OF THE SIMULATED TRAFFIC ACCIDENT INVESTIGATION

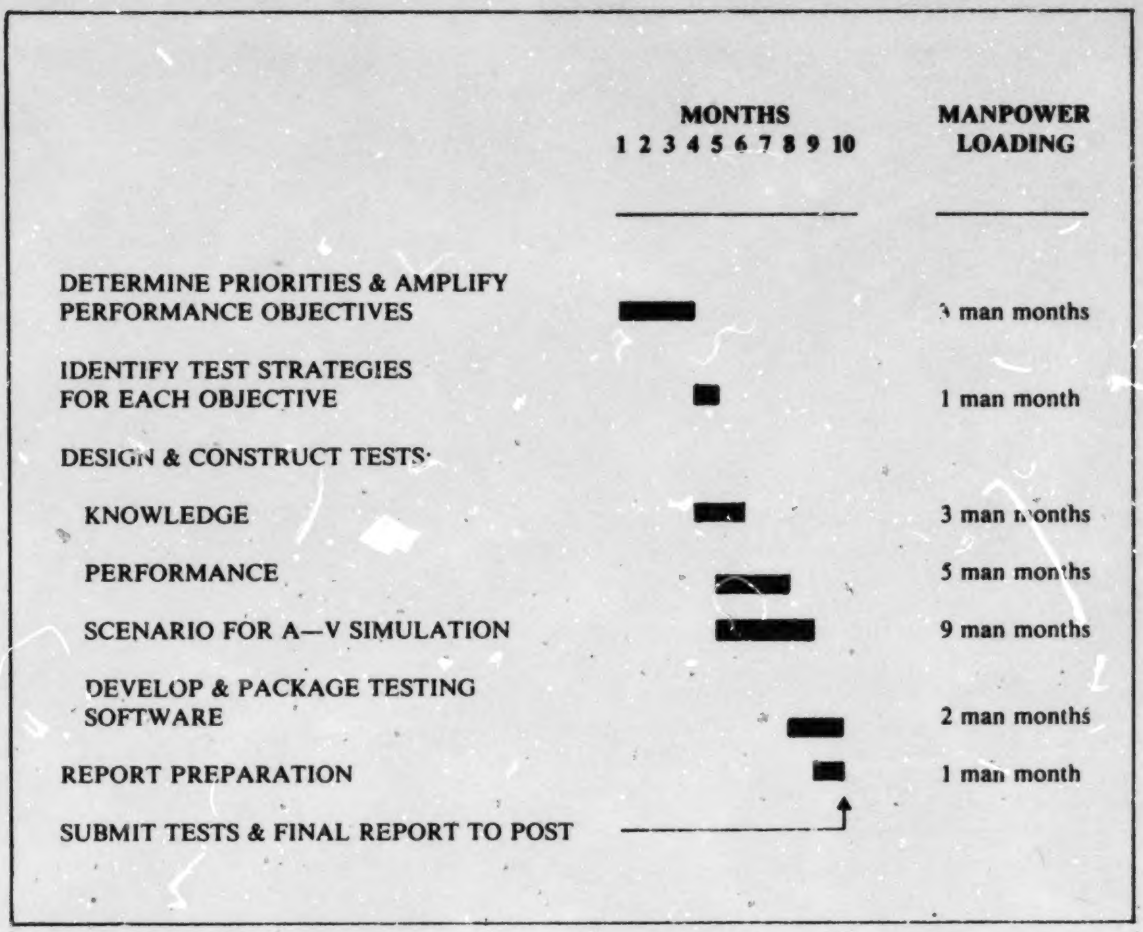
STEPS PERFORMED	STEPS OMITTED
Quick response	•
•	Request for additional information
•	•
•	•
Used emergency lights and siren if needed	Proper parking of police vehicle at scene
Attended injured	•
Called for ambulance and wrecker	•
•	•
Checked cars and made measurements	Checking of identification
•	Separate questioning of drivers
•	•
•	•
Recorded names and addresses of persons involved	•
•	Determination of destinations of wrecker and ambulance
•	•
•	•
Kept onlookers moving	•
•	•
•	Taking detailed notes
Controlled situation	•
•	Treating all with courtesy
•	
Maintained professional image	

RECOMMENDED DELIVERABLES

The following deliverables should be provided to POST at the conclusion of the contract.

- 1. A final report describing the test development process and products.
- 2. Camera ready copies of Performance Tests, Knowledge Tests, and A—V Simulation Tests, to include:
  - a. Tests
  - b. Administrator score sheets and examinee answer sheets.
  - c. Training materials and instructions for test administrators.
  - d. Instructions for examiners.
  - e. Scoring instructions.
  - f. A[illegible] scenarios for A—V simulations.

RECOMMENDED SCHEDULE OF ACTIVITIES & MANPOWER LOADING REQUIREMENTS



STAFFING REQUIREMENTS

Test development for POST should be undertaken by an organization with documentable competence and experience in the following areas:

1. The actual development and implementation of test instruments and programs to assess job knowledge and performance. Applications should include:
   a. Job sample test development.
   b. Criterion-referenced knowledge test development.
   c. Conceptual analysis of the relationships between performance tests and knowledge tests.
2. The development and implementation of performance-oriented job training programs which integrate instruction with tests for assessing course progress and terminal performance. This should include the derivation of both *training* and *testing* content from training objectives.
3. Development of A—V simulations for depicting performance in job situations.
4. Working directly with law enforcement agencies to develop training/testing systems for police officers.